SPECIFICATION:

Page 4, sixth paragraph, replace with the following new paragraph:

The method of ~~Moeslund et. al [1]~~ Moeslund T.B., et al., 4th IEEE Int. Conf. Automatic Face and Gesture Rec., 2000, p.362-367, utilizes color segmentation of the hand and the head in two cameras. This approach fails if the segments of head and hand come too close to each other.

Page 4, last paragraph (extends to page 5), replace with the following new paragraph:

The methods of ~~Goncalves et. al [2]~~ Goncalves L., et al., Proc. International Conference on Computer Vision, 1995, p. 764-770, and ~~Filova et. al [3]~~ Filova V., et al., Machine Vision and Application, 1998, 10: p. 223-231, perform model based tracking of a human arm in a single camera view. This approach obtains 3D information even in a single camera image, however, model based tracking as described in their paper is computationally extremely expensive and not suitable for practical application. Furthermore, the operating conditions are very constrained requiring the person whose arm is tracked to assume a very specific pose with respect to the camera.

Page 5, second paragraph, replace with the following new paragraph:

The method of ~~Wu et. al [4]~~ Wu A., et al., 4th IEEE Int. Conf. Automatic Face and Gesture Rec., 2000, p. 536-542, is also a model based approach and requires the detection of a users elbow and shoulder, which is difficult to perform outside of very constrained environments. More specifically, their method is based on skin color cues and implicitly assumes that the user, whose arm is being tracked, wears short-sleeved shirts, thus very much limiting the domain in which their method would be useful.

Page 5, third paragraph, replace with the following new paragraph:

The method of ~~Ahmad [5]~~ <u>Ahmad S., A Usable Real-Time 3D Hand Tracker, IEEE</u> <u>Asian Conference, 1994,</u> is able to track a human hand held between a camera and a table, where the camera is pointed at the table with the imaging sensor parallel to the table surface. Their method is very specific in that it is only usable in a situation where the user, whose hand is being tracked, is sitting at a table with his hand at a particular location held in a particular ~~pose and thus~~ <u>pose, and thus</u> lacks generality.

Page 13, first paragraph, replace with the following new paragraph:

The final measure of confidence of observing a hand at location s in an image is given by a combination of the color and the motion cue at that location. The proposed method and apparatus is hence tuned towards tracking and locating targets that are moving and have a color according to the target color histogram. However, as illustrated in Fig. 5 on the example of a moving black rectangle, the color confidence image and the motion confidence image have to be post-processed before they can be utilized. The rectangle is shown as 650 at time t-dt in an image 600 captured by a camera and as [[650c]] <u>650B</u> captured at a later ~~time time t~~ <u>time, time t,</u> in an image 610 by the same camera. While the motion cue function $M(s,t)$ (630) for this target tends to have high responses at the edges 631 and 632 of the target, the color cue function 620 tends to have high responses at the interior of the target 621. A logical or arithmetic combination of the two cues shown in image 640 only shows a response where both 620 and 630 show a response, i.e., where the cues intersect (at 641) if overlaid on top of each other.